IPST Technical Paper Series Number 569

Gene Expression Under Water Deficit in Loblolly Pine (*Pinus taeda* L.): Isolation and Characterization of cDNA Clones

S. Chang, J.D. Puryear, M.A. Dilip Dias, E.A. Funkhouser, R.J. Newton, and J. Cairney

April 1995

Submitted to
Physiologia Plantarum

Gene expression under water deficit in loblolly pine (*Pinus taeda* L.): Isolation and characterization of cDNA clones

Shujun Chang, Jeffrey D. Puryear, M. A. Dilip Dias, Edward A. Funkhouser, Ronald J. Newton, and John Cairney

Chang, S., Puryear, J. D., Dias, M. A. D. L., Funkhouser, E. A., Newton, R. J., & Cairney, J. 1995. Gene expression under water deficit in loblolly pine (*Pinus taeda* L.): Isolation and characterization of cDNA clones. - Physiol. Plant.

Poly(A^+) RNA was extracted from roots of loblolly pine (*Pinus taeda* L.) seedlings subjected to gradual and prolonged water deficit. This RNA was used to construct a cDNA library. A number of cDNA clones were isolated whose expression was induced by water deficit. Four of these cDNA clones, designated pLP2, pLP3, pLP4, and pLP5, were characterized further. Each of these pine genes has unique characteristics either in sequence or pattern of expression. The protein encoded by pLP2 shows 91% identity to *S*-adenosylmethionine synthetase from a range of plants. The protein encoded by pLP3 is similar to a tomato protein induced by water deficit and during fruit ripening, but the pine protein possesses a unique 34-amino-acid region near the amino terminal. Similarly, the LP4 protein is homologous to stellacyanin, a copper-binding protein in the Japanese lacquer tree, but possesses a proline/serine-rich carboxy terminus not found in other plants. Clone pLP5 encodes a novel glycine-rich protein with homology to silk fibroin and the rat chondroitin core protein, but the putative pine protein is distinct from previously characterized glycine-rich proteins. Transcript levels of the four genes rose under

moderate water deficit stress and then declined as stress became severe (one month without water), with the exception of pLP5 mRNA, which remained at elevated levels even under severe stress. The possible roles of the encoded proteins in cell wall reinforcement are discussed.

Key words - cDNA clones, loblolly pine, *Pinus taeda*, RNA isolation, stress, water deficit.

S. Chang and J. Cairney, Institute of Paper Science and Technology, 500 10th Street N.W., Atlanta, Georgia 30318, J. D. Puryear, M.A. Dias and R. J. Newton, Department of Forest Science and E. A. Funkhouser, Department of Biochemistry/Biophysics, Texas Agricultural Experiment Station, Texas A&M University System, College Station, Texas 77843.

Introduction

Water deficit in conifers is a major problem in commercial tree production, resulting in the loss of millions of dollars annually due to seedling mortality and impaired growth (Zahner and Donnelly 1967). Water deficit stress in conifers has been the focus of considerable attention. Research on water deficit stress in pines has indicated physiological mechanisms of adaptation (Pallardy 1981 , Newton et al. 1990) and the genetic basis of tolerance through classical breeding programs (van Buijtenen et al. 1976). Understanding of the distinctive physiology of conifers, particularly under stress, would be enhanced by a determination of gene expression under water deficit and by the isolation of novel stress-inducible pine genes.

Much recent research on plant responses to water deficit has been directed toward the isolation and identification of water deficit stress-inducible genes in the belief that such knowledge will indicate adaptive or defensive mechanisms (Skriver and Mundy 1990, Bray 1991, 1993). A variety of experimental conditions have been used in these studies, which usually exposed plants to rapid dehydration regimes. Clones isolated in these studies, through differential screening of

cDNA libraries, often show homologies to genes associated with plant responses to cell damage (Singh et al. 1989, Borkird et al. 1991, Downing et al. 1992), to genes whose products are involved in cell wall reinforcement (Gomez et al. 1988, Showalter et al. 1992, Keller 1993) and osmotic regulation (Guerrero et al. 1990, McCue and Hanson 1992, Niu et al. 1993, Verbruggen et al. 1993), and to genes expressed during the desiccation phase of embryo maturation (Mundy and Chua 1988, Close et al. 1989, Dure et al. 1989, Bray 1991).

The recalcitrance of pine tissue to biochemical anslysis has hampered similar research in conifers. Recent improvements in the technique of RNA isolation from pine (Chang et al. 1993) have allowed us to demonstrate changes in gene expression in pine seedlings subjected to water deficit stress (Funkhouser et al. 1993). In these experiments, stress was imposed by withholding irrigation from seedlings in a greenhouse to simulate periods of progressive drought. Here we described the cDNA cloning, sequencing, and expression analysis of four loblolly pine (*Pinus taeda* L.) genes induced by water deficit stress. Similarities to stress-inducible proteins from other plants are evident; however, these putative pine polypeptides possess unusual sequence features and their cognate genes displayed distinctive patterns of expression.

Abbreviations - GRPs, glycine-rich proteins; CTAB, hexadecyltrimethylammonium bromide; HPRGPs, hydroxyproline-rich glycoproteins; *lea*, late embryogenesis abundant; PG19, rat proteoglycan core protein 19; pI, isoelectric point; *rab*, responsive to ABA; SAM, *S*-adenosylmethionine; Sc, stellacyanin.

Materials and methods

Plant material and water deficit stress treatments

P. taeda L. (loblolly pine) seedlings were full-sibling, resulting from the cross of S6PT2 and S6PT3 sources from east Texas. The seeds were sown in 5-liter cylindrical containers filled with a fritted clay medium adapted for pine seedlings (Meier et al. 1992) and were watered thrice weekly until seedlings were established. The seedlings were then irrigated once weekly with nutrient solution as they grew to heights of 30 to 45 cm within 8 to 13 months in the greenhouse, where all experiments were completed.

In the water deficit stress experiments, seedlings were first acclimatized by irrigating daily with reverse osmosis water for 7 days. The seedlings were randomly separated into 7 groups, and water was then withheld from the first seedling group, while the other groups continued to receive daily irrigation. After several days, water was withheld from a second group, and this process continued until groups of seedlings that had been deprived of irrigation for 5, 11, 17, 20, 25, or 29 days were obtained. A group of control plants was irrigated daily. Greenhouse temperature ranged between 20 and 40°C. Plants were harvested predawn on the same day. Seedling predawn water potentials were determined using medium-age, fully-expanded needle fascicles in a pressure chamber by methods described previously (Meier et al. 1992). Immediately after a fascicle was taken for water potential measurement, each seedling was harvested: needles, stems and roots were quickly isolated and frozen separately by immersion in liquid nitrogen and stored at -80°C. RNA was later extracted from individual plants that had attained targeted water potentials representing a broad range of stress severities.

Molecular methods and cDNA library construction

The entire root tissue of an 8-month-old seedling with a water potential of -1.1 MPa was chosen for cDNA library construction. Total RNA was extracted as described below. Poly(A^+) RNA was isolated from total RNA using the PolyA Tract magnetic sphere system (Promega, Madison, WI). A cDNA library was prepared using the l-ZAP cDNA synthesis kit (Stratagene, La Jolla, CA) according to the manufacturer's recommendations. Poly(A^+) RNA from seedlings with water potentials of -0.4 MPa and -1.3 MPa was used to make control and stress probes for differential screening.

RNA isolation and northern blot hybridization

All needles, stems, or roots of each selected seedling were ground separately in liquid nitrogen. Total RNA was then extracted using a method developed in this laboratory (Chang et al. 1993). Equal 15 μg amounts of total RNA per sample (determined spectrophotometrically) were separated on a 1.2% agarose gel using formaldehyde and formamide as denaturants. Equal loading of each sample was checked by ethidium bromide staining. The RNA was blotted on Hybond-N^+ membranes (Amersham, Arlington Heights, IL).

DNA sequencing and homology comparison

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al. 1977). Standard molecular methods were used in other cases (Sambrook et al. 1989). For comparing the putative proteins encoded by these pine cDNA clones with similar proteins from other organisms, sequences were aligned using the LaserGene program (DNASTAR, Madison, WI).

Results

Our experimental design allowed a single-time harvest of plants that had experienced a gradual dehydration and now represented a broad range of stress levels, from a nonstressed mean of -0.37 MPa found with well-irrigated seedlings down to a mean of -2.23 MPa with the 29-day treatment (Fig. 1). Relatively few studies have examined gene expression in roots, despite observation of root growth in response to water deficit (Robertson et al. 1990) and evidence that stress signals may emanate from this organ (Davies and Zhang 1991, Tardieu et al. 1992). For these reasons, a cDNA library ($7.2x10^6$ pfu) was constructed from poly(A^+) RNA isolated from the roots of an 8-month-old, water deficit-stressed pine seedling. Water deficit stress-regulated genes were isolated by differential screening of about 15,000 independent plaques. From 28 putative water deficit stress-responsive signals identified in a primary screening, 15 were confirmed by Northern analysis. Sequences were determined for four distinct clones, which contain cDNA molecules whose sizes are similar to those of the mRNA to which they hybridized (Tab. 1). The patterns of expression, sequence homologies and characteristic features of each clone are considered below.

The cDNA of pLP2 hybridized to a single mRNA of approximately 1.6kb. The gene corresponding to pLP2 was water deficit-inducible in all organs, although the highest levels of expression were found in the stems (Fig. 2A). Expression rose as plants dehydrated, but declined as stress became severe, 25 to 29 days without irrigation (Fig. 2A, lanes d and e). Clone pLP2 contains a cDNA of 1485 nucleotides, similar in size to the mRNA, and has a long open reading frame, beginning with ATG, which could specify a protein of 393 amino acids (Fig. 3). A 5' untranslated leader of at least 100 nucleotides is inferred. The clone ends in a polyA tail, and there is a canonical polyadenylation signal sequence centered 118 nucleotides upstream of the tail. Plant polyadenylation sequences vary in position and sequence from those of mammals

(Joshi 1987, Mogen et al. 1990), so alternative cryptic signals may be involved in polyadenylation.

The putative protein encoded by this pLP2 clone is virtually identical to *S*-adenosylmethionine (SAM) synthetase from a number of plant species (Fig. 4, Peleman et al. 1989, Larsen and Woodson 1991, Kawalleck et al. 1992). SAM synthetase catalyzes the biosynthesis of SAM from methionine and ATP (Tabor and Tabor 1984). SAM is a cofactor in numerous biochemical reactions, acting as a methyl donor to proteins, lipids, polysaccharides, nucleic acids (Tabor and Tabor 1984), and intermediates in lignin synthesis (Sederoff and Chang 1991). SAM is an intermediate in the synthesis of ethylene (Kende 1993).

The gene corresponding to clone pLP3 was water deficit-inducible and was expressed predominantly in roots (Fig. 2B). While roots produced much higher absolute levels of mRNA than did stems and needles, the pattern of induction was similar. The mRNA is approximately 1.0 kb (Tab. 1), while the cDNA is 778 nucleotides in length (Fig. 5). The mRNA of pLP3 could encode a 153-amino-acid polypeptide having 49% identity and 56% similarity to the putative tomato protein TMA SN1 (Fig. 6, Iusem et al. 1993). Interestingly, the putative loblolly pine protein possesses a region of 34 amino acids (Pro24 to Ala58) which is absent from the tomato protein. If this region is excluded and the match reconsidered, the identity rises to 62%. The tomato protein is water deficit-inducible and is expressed during fruit ripening (Iusem et al. 1993). Cell fractionation experiments which suggest that TMA SN1 is located principally in the nucleus, along with the protein's known basic nature (pI 12.9), led Iusem et al. (1993) to suggest that TMA SN1 may be a chromatin-associated protein. Protein LP3 is more acidic (pI 6.09, Tab. 1), mainly due to the additional peptide sequence. No cell localization data are available for the pine protein.

The pLP4 gene was water deficit-inducible in needles, stems and roots (Fig. 2C). Expression of a 1.2 kb mRNA diminished slightly under extreme stress, but more notable was the appearance of a second mRNA species of 1.57 kb as stress increased. It is assumed that the

cDNA corresponds to the more prominent, 1.2 kb RNA, since the 1.57 kb band was a relatively minor species in roots; identification must be confirmed. The cDNA insert of clone pLP4 at 745 nucleotides is much shorter than either of the hybridizing RNAs (Tab. 1). DNA sequencing showed an open reading frame beginning with histidine at the very start of the cDNA (Fig. 7). These two results suggest that pLP4 lacks information from the 5' end of the transcript. Tentative identification, however, is possible: the predicted polypeptide from pLP4 is similar to a number of Type I ("blue") copper-containing glycoproteins from the Japanese lacquer tree (*Rhus vernecifera*), cucumber, and *Arabidopsis* (Fig. 8, Fields et al. 1991, Murata et al. 1982, van Gysel et al. 1992).

A comparison of the protein putatively encoded by pLP4 with stellacyanin (Sc), a glycoprotein from the Japanese lacquer tree, reveals both similarities and differences. From amino acid 1 to 79, homology to the Sc protein is strong (43% identity, 61% similarity, Fig. 8). Residues corresponding to the copper-binding sites (His12, Cys53, Cys58, Met63), disulfyl bonding sites (Cys25, Cys59), and the sequences around them compare closely (Fig. 8, Fields et al. 1991). However, protein LP4 has an additional 100 amino acids beyond the region of homology to Sc (position 79-179, Fig. 8). In this respect, the LP4 protein resembles the *Arabidopsis* polypeptide, which has a carboxy-terminal extension. This part of the LP4 sequence contains 29% Ser, 23% Pro, and 20% Thr. The disproportionate representation of these amino acids in the terminal 100 residues of the protein is evident when one considers that Ser, Pro, and Thr comprise 5%, 1%, and 12% respectively of the preceding 79 amino acids. Over the 179 residues of sequence available, Ser, Pro and Thr form 18%, 13%, and 12% of residues respectively.

Between residues 79 and 179, several Pro-Ser motifs are evident: PSPS is repeated five times, PSPSP three times, SSPP two times. The PSPSP motif is confined to the region of the polypeptide unique to pine (amino acids 81-110), a region composed almost exclusively of Pro and Ser residues. The motifs SPPP, STTT, SAAA, and SLLL each appear once in the terminal 100 amino acids of LP4, but they are absent from the *Arabidopsis* protein (Fig. 8). It is

noteworthy in this regard that certain Tyr residues, at positions 100, 106, 116, 122, and 132, are spaced 5, 9, 5, and 9 amino acid residues apart. Keller (1993) suggested that the defined spacing of Tyr could control the crosslinking and thus the porosity of the cell wall. Many cell wall proteins are pro rich (Showalter and Varner 1989, Varner and Lin 1989), however, we have no data as yet on the localization of these proteins.

Clone pLP5 represents a water deficit-inducible gene which was expressed almost exclusively in roots; very small amounts of transcript could be seen in stems and none in needles (Fig. 2D). Unlike most genes discussed thus far, the expression of LP5 did not seem to diminish even under severe stress. At 987 nucleotides, the cDNA clone pLP5 is very similar in size to the mRNA (Tab. 1). There are two ATGs near the 5'end of the molecule (Fig.9); the context of the ATG (nucleotides 99-101)is very similar to the consensus eukaryotic translational start sequence (Kozak 1989) while ATG (nucleotide 65-67) shows little similarity to the consensus. As yet we have no information as to which is used. The open reading frame commencing at nucleotide 99 could encode a polypeptede of 194 amino acids (Fig. 9) which is very rich in Gly (44%) and Ser (20%). The accumulation of glycine-rich proteins (GRPs) in stressed plants has been documented in many cases (Condit and Meagher 1986, Gomez et al. 1988, Keller et al. 1988, 1989, 1993, Showalter et al. 1992). These proteins are expressed in vascular systems and are presumed to be located in the cell wall (Keller et al. 1988, 1989, 1993). The putative LP5 protein, however, shows the greatest similarity to silk fibroin, and peptidoglycan core protein constituents of the extracellular matrix are thought to fulfill a structural role in the cell (Bourdon et al. 1986, Parthasarathy and Tanzer 1987).

Discussion

None of the clones reported in this study or recently isolated by us (nonpublished data) show sequence homology with the dehydrin, *lea*, or ABA-responsive (*rab*) genes, which have been

isolated in similar studies (Galau et al. 1986, Baker et al. 1988, Mundy and Chua 1988, Close et al. 1989). As noted earlier, in many of these other studies, imposition of stress was rapid and RNA isolation occured soon after the commencement of stress, usually within a few hours (Close et al 1989, Guerrero et al 1990, Skriver and Mundy 1990). Seedlings in the present study were allowed a more gradual dehydration in a rooting medium that was chosen for its soil-like drying characteristics, which would simulate water deficit stress conditions experienced by trees in their natural habitat. The library was constructed from seedlings that had experienced stress for about two weeks. The focus of these experiments was therefore towards genes whose expression remains at an elevated level over extended periods of gradually increasing stress.

Recently, Leone et al (1994) showed that in potato cell suspension cultures, rapid versus gradual imposition of water stress induced different sets of polypeptides. Those proetin induced by abrupt stress treatment were also ABA-inducible; those induced in gradually stressed cells were not ABA-inducible.

In our study, gene expression was monitered over a rage of stresses and maximal steady state RNA levels were observed between -0.8 MPa and -0.13 MPa (Fig. 2A-D). This is the stress range at which library construction and clone selection took place. Under more severe stress, RNA levels, in most cases, declined. Expression may therefore be regarded as transient, though stress levels change over the course of the experiment (as they do in the field) so the transient nature of pLP 2 ,3, 4 and 5 gene expression should be distinguished from the rapid bursts of gene expression monitered by other workers (cf. Leone et al. 1994).

The trend appears to be toward the isolation of genes whose functions may be in cell wall reinforcement. Such an activity would certainly be a continuing need for turgor maintenance in a plant subjected to prolonged and severe stress. Recently, induction of SAM synthetase gene expression by fungal elicitors has been shown in parsley (Kawalleck et al. 1992), and Cruz et al. (1992) demonstrated increased lignification of xylem cell walls in maize subjected to water

deficit stress. These observations seem consistent and compatible with our own demonstration of SAM synthetase gene induction in response to an environmental stress.

The LP4 protein has a proline/serine rich domain, a feature found in many cell was proteins (Showalter and Varner 1989, Varner and Lin 1989). The amino-terminal portion of LP4 is homologous to Sc, a Type I "blue" copper-binding glycoprotein from the Japanese lacquer tree. Notable among the copper-containing proteins are laccase and peroxidase, both of which have been implicated in the polymerization of monolignols into lignin in the cell wall (O'Malley et al. 1993, Olsen and Varner 1993).

The striking similarities of LP5 to specialized mammalian structural proteins suggest that LP5 may play a similar role to mediate the elasticity and the strength of the wall (Keller 1993). Cell wall proteins expressed at the sites of lateral root formation and associated with changes in root hair morphology have been documented in plants (Keller et al. 1989, Schnall and Quatrano 1992). The preferential expression of the pLP5 gene in roots may be related to such processes.

The polypeptide encoded by the pLP3 gene is similar to the tomato protein TMA SN1, which is water deficit-inducible and is also expressed during fruit ripening. Iusem et al. (1993) reported that this protein is located primarily in the nucleus and noted similarities to eukaryotic chromosomal proteins. DNA supercoiling and chromatin structure have considerable influence on gene expression (Pruss and Drlica 1989, Zlatanova 1990, Felsenfeld 1992) and in bacteria, environmentally induced changes in DNA supercoiling may be mediated by basic histone-like proteins (Hulton et al. 1990, Goransson et al. 1990). Preliminary results suggest that pLP3 is present as a family of genes, and two loblolly pine genomic clones have been isolated (non-published data). This contrasts with the situation in tomato where TMA SN1 is present as a single gene (Iusem et al. 1993).

In summary, cDNA clones of genes with a variety of possible functions have been isolated from loblolly pine seedlings subjected to water deficit stress. These are the first water deficit-

inducible clones isolated from gymnosperms using elevated expression as a selection criterion. The predicted polypeptides are distinctive, but homologies with previously characterized proteins suggest that some of the pine proteins may fulfill a structural role either directly or through participating in the synthesis of cell wall components. Their sequence similarities and patterns of expression suggest that the proteins may participate in the adaptation of cells to extended periods of stress.

Acknowledgements - The authors thank Martha Swize for technical assistance and J. P. van Buijtenen, Department of Forest Science, Texas A&M University, for providing loblolly pine seeds. This work was supported by grants from the Temple Inland Foundation, Diboll, Texas; the Texas Agricultural Experiment Station; and US-Israel BARD Grant No. US-2212-92. TAES Technical Article No. 31545.

References

Baker, J., Steele, C. & Dure III, L. 1988. Sequence and characterization of 6 Lea proteins and their genes from cotton. - Plant Mol. Biol. 11: 277-291.

Borkird, C., Simeons, C., Villarroel, R. & van Montagu, M. 1991. Gene expression associated with water stress adaptations of rice cells and identification of two genes as hsp 70 and ubiquitin. - Physiol. Plant. 82: 449-457.

Bourdon, M. A. , Shiga, M. & Ruoslahti, E. 1986. Identification from cDNA of the precursor form of a chondroitin sulfate proteoglycan core protein. - J. Biol. Chem. 261: 12534-12537.

Bray, E. A. 1991. Regulation of gene expression by endogenous ABA during drought stress. - *In* Abscisic Acid: Physiology and Biochemistry (W. J. Davies and H. G. Jones, eds.), pp. 81-98. Bios Scientific Publishers, Oxford, UK.

- 1993. Molecular responses to water deficit. - Plant Physiol. 103: 1035-1040.

Chang, S., Puryear, J. & Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. - Plant Mol. Biol. Rep. 11: 114-117.

Close, T. J., Kortt, A. A. & Chandler, P. M. 1989. A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. - Plant Mol. Biol. 13: 95-108.

Condit, C. M. & Meagher, R. B. 1986. A gene encoding a novel glycine-rich structural protein of petunia. - Nature 323: 178-181.

Cruz, R. T., Jordan, W. R. & Drew, M. C. 1992. Structural changes and associated reduction of hydraulic conductance in roots of *Sorghum bicolor* L. following exposure to water deficit. - Plant Physiol. 99: 203-212.

Davies, W. J. & Zhang, J. 1991. Root signals and the regulation of growth and development of plants in drying soil. - Annu. Rev. Plant Physiol. Plant Mol. Biol. 42: 55-76.

Downing, W. L., Mauxion, R., Fauvarque, M-O., Reviron, M-P., de Vienne, D., Vartanian, N. & Giraudat, J. 1992. A *Brassica napus* transcript encoding a protein related to the Kunitz

protease inhibitor family accumulates upon water stress in leaves, not in seeds. - Plant J. 2: 685-693.

Dure III, L., Crouch, M., Harada, J., Ho, T. H. D., Mundy, J., Quatrano, R., Thomas, T. & Sung, Z. R. 1989. Common amino acid sequence domains among the LEA proteins of higher plants. - Plant Mol. Biol. 12: 475-486.

Esau, K. 1965. Plant Anatomy. John Wiley & Sons, New York.

Felsenfeld, G. 1992. Chromatin as an essential part of the transcriptional mechanism. - Nature 355: 219-224.

Fields, B. A., Guss, J. M. & Freeman, H. C. 1991. Three-dimensional model for stellacyanin, a "blue" copper-protein. - J. Mol. Biol. 222: 1053-1065.

Funkhouser, E. A., Cairney, J., Chang, S., Dias, M. A. D. & Newton, R. J. 1993. Cellular and molecular responses to water deficit stress in woody plants. - *In* Handbook of Plant and Crop Stress (M. Pessarakli, ed.), pp. 347-362. Marcel Dekker, New York.

Galau, G. A., Hughes, D. W. & Dure III, L. 1986. Abscisic acid induction of cloned cotton late embryogenesis abundant (LEA) mRNAs. - Plant Mol. Biol. 7: 155-170 .

Gomez, J., Sanchez-Martinez, D., Stiefel, V., Rigau, J., Puigdomenech P. & Pages, M. 1988. A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. - Nature 334: 262-264.

Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. & Uhlin, B. E. 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. - Nature 344: 682-685.

Guerrero, F. D., Jones, J. T. & Mullet, J. E. 1990. Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes. - Plant Mol. Biol. 15: 262-264.

Hulton, C. S. J., Seirafi, A., Hinton, J. C. D., Sidebotham, J. M., Waddell, L., Pavitt, G. D., Owen-Hughes, T., Spassky, A., Buc, H. & Higgins, C. F. 1990. Histone-like protein H1 (H-NS), DNA supercoiling and gene expression in bacteria. - Cell 63: 631-642.

Iusem, D. N., Bartholomew, D. M., Hitz, W. D. & Scolnik, P. A. 1993. Tomato (*Lycopersicon esculentum*) transcript induced by water deficit and ripening. - Plant Physiol. 102: 1353-1354.

Joshi, C. P. 1987. Putative polyadenylation signals in nuclear genes of higher plants: a compilation analysis. - Nucl. Acids Res. 15: 9627-9640.

Kawalleck, P., Plesch, G., Hahlbrock, K. & Somssich, I. E. 1992. Induction by fungal elicitor of S-adenosyl-L-methionine synthetase and S-adenosyl-L-homocysteine hydrolase mRNAs in cultured cells and leaves of *Petroselinum crispum*. - Proc. Natl. Acad. Sci. USA 89: 4713-4717.

Keller, B. 1993. Structural cell wall proteins. - Plant Physiol. 101: 1127-1130.

- , Sauer, N. & Lamb, C. J. 1988. Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. - EMBO J. 7: 3625-3633.

- , Templeton, M. D. & Lamb, C. J. 1989. Specific location of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. - Proc.Natl. Acad.Sci. USA 86: 1529-1533.

Kende, H. 1993. Ethylene biosynthesis. - Annu. Rev. Plant Physiol. Plant Mol. Biol. 44: 283-307.

Larsen, P. B. & Woodson, W. R. 1991. Cloning and nucleotide sequence of an S-adenosylmethionine synthetase cDNA from carnation. - Plant Physiol. 96: 997-999.

Lei, M. & Wu, R. 1991. A novel glycine-rich cell wall protein gene in rice. - Plant Mol. Biol. 16: 187-198.

Leone, A., Costa, H., Tucci, M. & Grillo, S. 1994. Comparative analysis of short- and long-term changes in gene expression caused by low water potential in potato (*Solanum tuberosum*) cell-suspension cultures. Plant Physiol. 106:703-712.

McCue, K. F. & Hanson, A. D. 1992. Salt-inducible betaine aldehyde dehydrogenase from sugar beet: cDNA cloning and expression. - Plant Mol. Biol. 18: 1-11.

Meier, C. E., Newton, R. J., Puryear, J. D. & Sen, S. 1992. Physiological responses of loblolly pine (*Pinus taeda* L.) seedlings to drought stress: osmotic adjustment and tissue elasticity. - J. Plant Physiol. 140: 754-760.

Mita, K., Ichimura, S., Zama, M. & James, T. C. 1988. Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA. - J. Mol. Biol. 203: 917-925.

Mogen, B. D., MacDonald, M. H., Graybosch, R. & Hunt, A. G. 1990. Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants. - Plant Cell 2: 1261-1272.

Mundy, J. & Chua, N-H. 1988. Abscisic acid and water-stress induce the expression of a novel rice gene. - EMBO J. 7: 2279-2286.

Murata, M., Begg, G. S., Lambrou, F., Leslie, B., Simpson, R. J., Freeman, H. C. & Morgan, F. J. 1982. Amino acid sequence of a basic blue protein from cucumber seedlings. - Proc. Natl. Acad. Sci. USA 79: 6434-6437.

Newton, R. J., Funkhouser, E. A., Fong, F. & Tauer, C. G. 1990. Molecular and physiological genetics of drought tolerance in forest species. - For. Ecol. Manage. 43: 225-250.

Niu, X., Zhu, F-K., Narasimhan, M. L., Bressan, R. A. & Hasegawa, P. M. 1993. Plasma-membrane H^+-ATPase gene expression is regulated by NaCl in cells of the halophyte *Atriplex nummularia* L. - Planta 190: 433-438.

Olsen, P. D. & Varner, J. E. 1993. Hydrogen peroxide and lignification. - Plant J. 4: 887-892.

O'Malley, D. M., Whetten, R., Bao, W., Chen, C-L. & Sederoff, R. R. 1993. The role of laccase in lignification. - Plant J. 4: 751-757.

Pallardy, S. G. 1981. Closely related woody plants. - *In* Water Deficits and Plant Growth (T. T. Kozlowski, ed.), Vol. 6, pp. 511-548. Academic Press, New York.

Parthasarathy, N. & Tanzer, M. L. 1987. Isolation and characterization of a low molecular weight chondroitin sulfate proteoglycan from rabbit skeletal muscle. - Biochemistry 26: 3149-3156.

Peleman, J., Boerjan, W., Engler, G., Seurinck, J., Botterman, J., Alliotte, T., van Montagu, M. & Inze D. 1989. Strong cellular preference in the expression of a housekeeping gene of *Arabidopsis thaliana* encoding S-adenosylmethionine synthetase. - Plant Cell 1: 81-93.

Pruss, G. & Drlica, K. 1989. DNA supercoiling and prokaryotic transcription. - Cell 56: 521-523.

Robertson, J. M., Hibick, K. T., Yeung, E. C. & Ried, D. M. 1990. Developmental responses to drought and abscisic acid in sunflower roots. I. Root growth, apical anatomy, and osmotic adjustment. - J. Exp. Bot. 41: 325-337.

Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, Ed. 2. - Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. - Proc. Natl. Acad. Sci. USA 74: 5463.

Schnall, J. A. & Quatrano, R. S. 1992. Abscisic acid elicits the water-stress response in root hairs of *Arabidopsis thaliana* . - Plant Physiol. 100: 216-218.

Sederoff, R. R. & Chang, H-M. 1991. Lignin biosynthesis. - *In* Wood Structure and Composition (M. Lewin and I. S. Goldstein, eds.), pp. 263-285. Marcel Dekker, New York.

Showalter, A. M. & Varner, J. E. 1989. Plant hydroxyproline-rich glycoproteins. - *In* The Biochemistry of Plants: A Comprehensive Treatise (A. Marcus, ed.), Vol. 15, pp. 485-520. Academic Press, New York.

- , Butt, A. D. & Kim, S. 1992. Molecular details of tomato extensin and glycine-rich protein gene expression. - Plant Mol. Biol. 19: 205-215.

Singh, N. K., Nelson, D. E., Kuhn, D., Hasegawa, P. M. & Bressan, R. A. 1989. Molecular cloning of osmotin and regulation of its expression by ABA and adaptation to low water potential. - Plant Physiol. 90: 1096-1101.

Skriver, K. & Mundy, J. 1990. Gene expression in response to abscisic acid and osmotic stress. - Plant Cell 2: 503-512.

Tabor, C. W. & Tabor, H. 1984. Methionine adenosyltransferase (S-adenosylmethionine synthetase) and S-adenosylmethionine decarboxylase. - Adv. Enzymol. 56: 251-282.

Tardieu, F., Zhang, F. & Davies, W. J. 1992. What information is conveyed by an ABA signal from maize roots in drying field soil? - Plant Cell Environ. 15: 185-191.

van Buijtenen, J. P., Bilan, M. V. & Zimmerman, R. H. 1976. Morphophysiological characteristics related to drought resistance in *Pinus taeda*. - *In* Tree Physiology and Yield Improvement (M. G. R. Cannell and F. T. Last, eds), pp. 349-359. Academic Press, New York.

van Gysel, A., Litiere, K., van Montagu, M. & Inze, D. 1992. - EMBL Data Library, Accession: S25555.

Varner, J. E. & Lin, L-S. 1989. Plant cell wall architecture. - Cell 56: 231-239.

Verbruggen, N., Villarroel, R. & van Montagu, M. 1993. Osmoregulation of a pyrroline-5-carboxylate reductase gene in *Arabidopsis thaliana*. - Plant Physiol. 103: 771-781.

Zahner, R. & Donnelly, J. R. 1967. Refining correlations of rainfall and radial growth in young red pine. - Ecology 48: 525-530.

Zlatanova, J. 1990. Histone H1 and the regulation of transcription of eukaryotic genes. - Trends Biochem. Sci. 15: 273-276.

Figure captions and legends

Fig. 1. Water potentials of medium-age fascicles of loblolly pine seedlings growing in fritted clay medium from which irrigation was withheld for the various time periods shown.

Fig. 2. Northern analyses of gene expresssion in 8-month-old loblolly pine seedlings subjected to water deficit, using probes A, pLP2; B, pLP3; C, pLP4; and D, pLP5. Seedlings were harvested and water potentials measured predawn on the same day; needles, stems and roots of each seedling were processed separately. Lane designations: a, -0.35 MPa, irrigated daily; b, -0.80 MPa, 11 days without irrigation; c, -1.30 MPa, 17 days without irrigation; d, -1.85 MPa, 25 days without irrigation; e, -2.30 MPa, 29 days without irrigation.

Fig. 3. cDNA nucleotide sequence of clone pLP2. The putative translation initiation and termination codons and polyadenylation signal are underlined. The EcoRI and XhoI cloning sites are included at each end of the cDNA as reference.

Fig. 4. Amino acid homology between LP2 and SAM synthetase from a number of plants (Peleman et al. 1989, Larsen and Woodson 1991, Kawalleck et al. 1992). The sequences were aligned using the LaserGene program (DNASTAR, Madison, WI). Identical residues are boxed, with dashes representing gaps.

Fig. 5. Nucleotide sequence of cDNA of clone pLP3. Underlined are the polyadenylation signal and putative translation initiation and termination codons. EcoRI and XhoI cloning sites are included at each end of the cDNA for reference.

Fig. 6. Sequence homology between LP3 and TMA SN1, a water deficit-inducible protein from tomato (Iusem et al. 1993). The LaserGene program (DNASTAR, Madison, WI) was used to align sequences. Identical residues are boxed, while dashes represent gaps.

Fig. 7. Nucleotide sequence of cDNA of clone pLP4. EcoRI and XhoI cloning sites are included at each end as reference. The putative translation termination codon and polyadenylation signal are underlined.

Fig. 8. Amino acid homology between LP4 and copper-containing proteins from Japanese lacquer tree (Fields et al. 1991), cucumber (Murata et al. 1982) and *Arabidopsis* (van Gysel et al. 1992). The sequences were aligned using the LaserGene program (DNASTAR, Madison,WI). Identical residues are boxed, with gaps represented by dashes.

Fig. 9. Clone pLP5 cDNA nucleotide sequence. Underlined are the 16-amino-acid tandem repeats, polyadenylation signal, and putative translation initiation and termination codons. The EcoRI and XhoI cloning sites are included at each end for reference.

Fig. 10. Sequence homology between LP5, silk fibroin (Mita et al. 1988) and rat proteoglycan core protein (Krusius and Ruoslahti 1986). The LaserGene program (DNASTAR, Madison, WI) was used to align sequences. Dashes represent gaps, while identical residues are boxed.

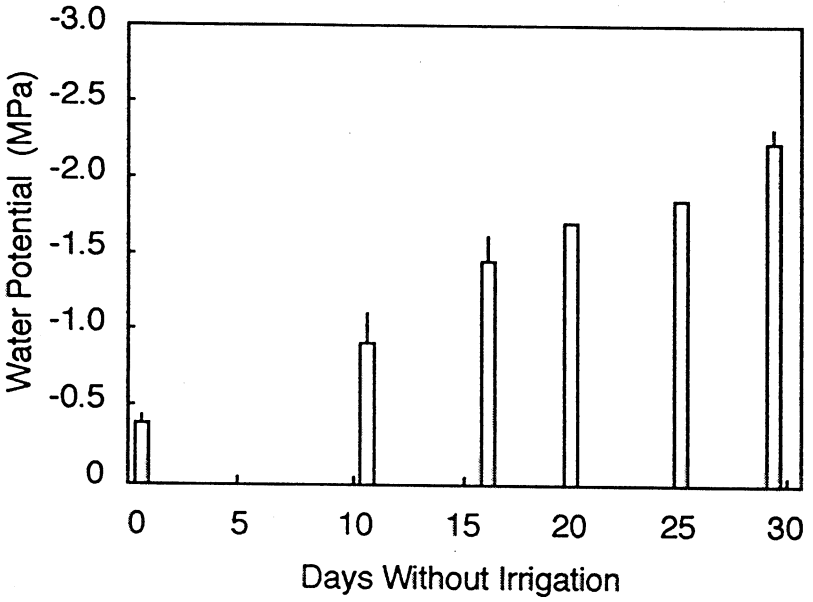
-3.0
-2.5
-2.0
-1.5
-1.0
-0.5
0
Water Potential (MPa)
0
5
10
15
20
25
30
Days Without Irrigation

Figure 1.

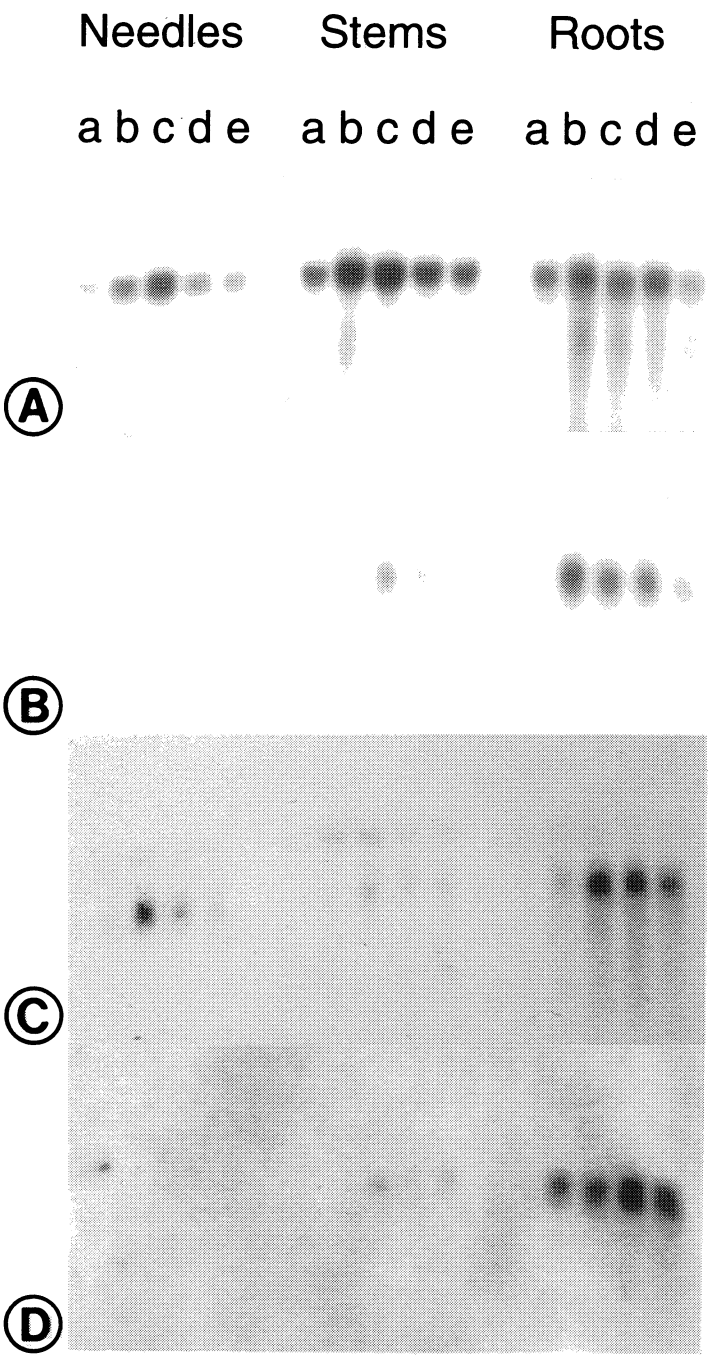

Figure 2.

```
GGCACGAGGCTCTATTAGCTCAGGCCTCTCTGCTGCTCCTTGGATTGCAATCTCGTTCTC      60
TGATTTGCCGTGCTGTTTGCTCTGCTCACTTCAGCCCAGATGGAGACCTTCTTGTTCACA     120
                                          M  E  T  F  L  F  T       7
TCAGAGTCTGTAAATGAGGGACACCCAGACAAACTCTGTGACCAGATTTCTGATGCAGTG     180
 S  E  S  V  N  E  G  H  P  D  K  L  C  D  Q  I  S  D  A  V      27
TTGGATGCATGCCTCACCCAGGACCCCGACAGCAAGGTAGCATGCGAGACTTGCACTAAA     240
 L  D  A  C  L  T  Q  D  P  D  S  K  V  A  C  E  T  C  T  K      47
ACGAACATGGTCATGGTTTTTGGTGAAATCACCACCAAGGCCGATGTCGATTATGAGCAG     300
 T  N  M  V  M  V  F  G  E  I  T  T  K  A  D  V  D  Y  E  Q      67
ATTGTTCGCAAGACCTGCAGGGAGATTGGTTTCATTTCTGACGATGTGGGTCTTGATGCT     360
 I  V  R  K  T  C  R  E  I  G  F  I  S  D  D  V  G  L  D  A      87
GATCACTGCAAAGTGCTGGTTAATATTGAACAGCAGAGCCCTGACATTGCCCAGGGAGTT     420
 D  H  C  K  V  L  V  N  I  E  Q  Q  S  P  D  I  A  Q  G  V     107
CATGGACACTTTACCAAGAGGCCTGAGGAGATTGGAGCTGGTGACCAGGGTCACATGTTT     480
 H  G  H  F  T  K  R  P  E  E  I  G  A  G  D  Q  G  H  M  F     127
GGATATGCAACTGATGAGACCCCTGAGCTGATGCCCCTGACCCATGTGCTGGCTACCAAG     540
 G  Y  A  T  D  E  T  P  E  L  M  P  L  T  H  V  L  A  T  K     147
CTGGGAGCGAAGCTCACCGAGGTCAGAAAGAATGGAACCTGCCCCTGGTTGAGGCCTGAT     600
 L  G  A  K  L  T  E  V  R  K  N  G  T  C  P  W  L  R  P  D     167
GGAAAAACCCAAGTGACTATTGAGTACCGAAACGAAGGGGGTGCCATGGTTCCTGAGCGG     660
 G  K  T  Q  V  T  I  E  Y  R  N  E  G  G  A  M  V  P  E  R     187
GTTCACACTGTTCTCATCTCCACTCAGCACGATGAGACAGTGACCAATGACCAGATTGCT     720
 V  H  T  V  L  I  S  T  Q  H  D  E  T  V  T  N  D  Q  I  A     207
GCAGATTTGAAGGAGCATGTAATAAAGCCGGTGATTCCTGAGAAGTACCTGGACGAGAAT     780
 A  D  L  K  E  H  V  I  K  P  V  I  P  E  K  Y  L  D  E  N     227
ACCATATTCCACTTGAACCCGTCTGGTCGATTCGTGATCGGAGGGCCTCATGGAGATGCA     840
 T  I  F  H  L  N  P  S  G  R  F  V  I  G  G  P  H  G  D  A     247
GGCCTCACCGGCAGGAAGATTATTATTGATACTTATGGAGGGTGGGGAGCTCATGGAGGA     900
 G  L  T  G  R  K  I  I  I  D  T  Y  G  G  W  G  A  H  G  G     267
GGTGCATTCTCTGGGAAGGATCCCACTAAGGTGGACCGAAGTGGGGCATACATAGTTAGA     960
 G  A  F  S  G  K  D  P  T  K  V  D  R  S  G  A  Y  I  V  R     287
CAGGCTGCCAAGAGCATTGTTGCAGCTGGACTTGCAAGGAGATGCCTTGTGCAGGTGTCT    1020
 Q  A  A  K  S  I  V  A  A  G  L  A  R  R  C  L  V  Q  V  S     307
TATGCCATCGGAGTGCCGGAGCCTCTGTCTATCTTTGTTGATTCGTATGGTACAGGGAGC    1080
 Y  A  I  G  V  P  E  P  L  S  I  F  V  D  S  Y  G  T  G  S     327
ATTCCAGACAAGGAAATTCTGGAGATAATTAAAGAGCACTTTGATTTCAGGCCTGGCATG    1140
 I  P  D  K  E  I  L  E  I  I  K  E  H  F  D  F  R  P  G  M     347
ATCACGATCAACCTTGATCTGAAGAGAGGAGGAAATGGAAGGTTCCAGAAGACGGCCGCC    1200
 I  T  I  N  L  D  L  K  R  G  G  N  G  R  F  Q  K  T  A  A     367
TATGGCCACTTTGGCAGGGATGATCCAGATTTCACCTGGGAGACTGTTAAGCCTCTTAAG    1260
 Y  G  H  F  G  R  D  D  P  D  F  T  W  E  T  V  K  P  L  K     387
TGGGAAAAGGCCCAAGCCTAAAATAGCAGCATTTTTCTTCATTGCCGATAACCTCATATC    1320
 W  E  K  A  Q  A  *                                            393
CCAGTACTGTTTTCTTTAGAAGTAGTAATAAAAATGTGTTAGTAATTGTGTGGTGTCGTC    1380
ACTTTAGCTTCGACCTGTGACTGTTTGGGTTATAGTGGCTGAATTTGTACTTCATATTTC    1440
ATCTATACTATTCGGTTCCATTCACATAAAAAAAAAAAAAAAAA                    1485
```

Figure 3.

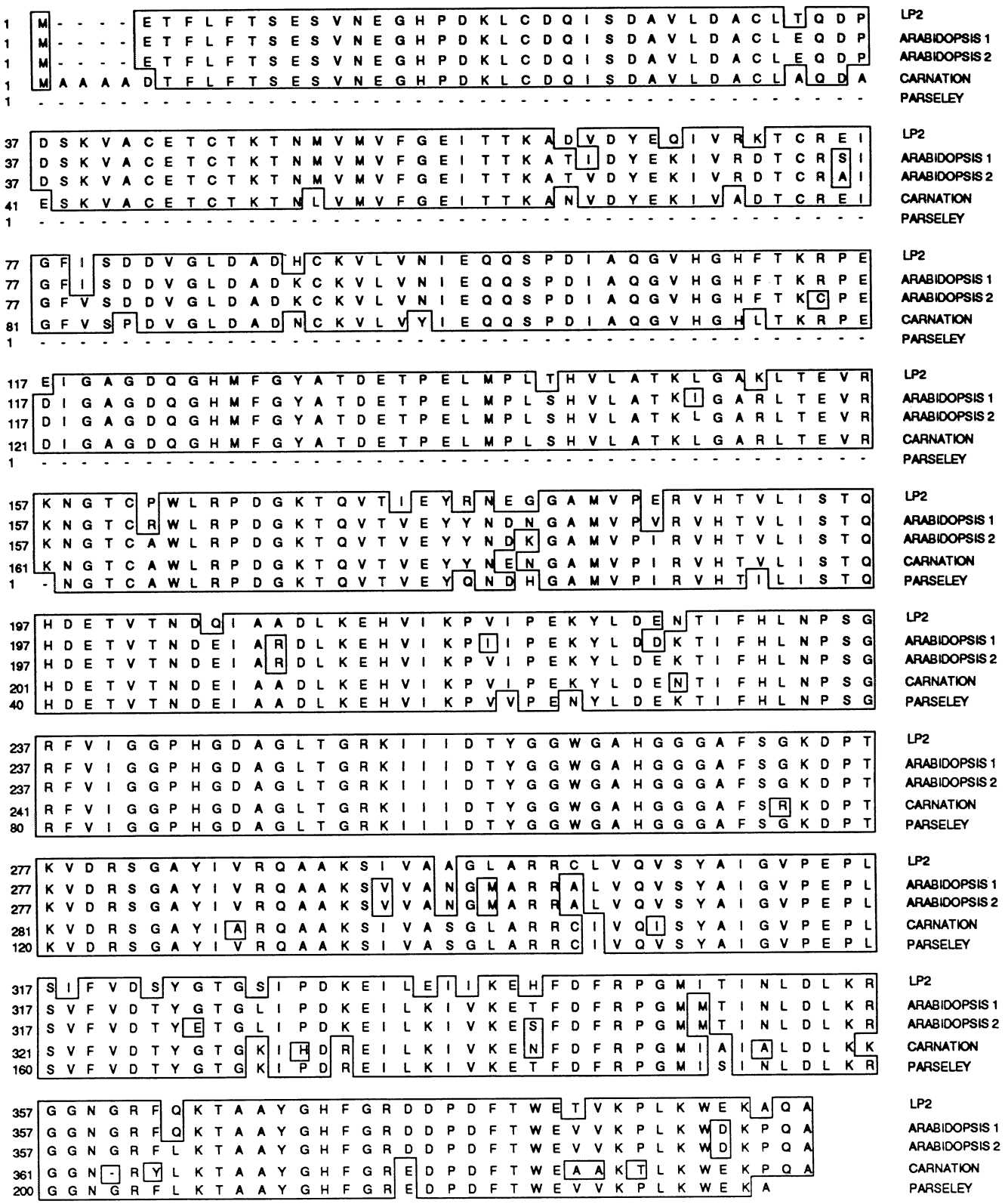

Figure 4.

```
GCAGGATTCGGCACGAGATTTACTTTGCTTCTTTGAATATTTGCTTGTTTTTGTGTGTAG    60
AAGATGTCTGAAGAAAAGCACCACCACCATCTGTTGCACCACAAGAAGGAAGATGAGAGC   120
    M  S  E  E  K  H  H  H  H  L  L  H  H  K  K  E  D  E  S     19
GAGAACGTGCCCTCGGAGGTTGTTTGCGCTGAGACCACCACAGCTTATGGCGATGAGGTG   180
 E  N  V  P  S  E  V  V  C  A  E  T  T  T  A  Y  G  D  E  V     39
ATCCAAAGCGCAGATGTGTACGCTGCTGGGGAGGTGAATGATGATAAGTTCGCGGAGTAC   240
 I  Q  S  A  D  V  Y  A  A  G  E  V  N  D  D  K  F  A  E  Y     59
GAGAAGGCGAGGAAGGAAGAGAAGCATCACAAGCATCTGGAGGAATTGGGTGGACTGGGA   300
 E  K  A  R  K  E  E  K  H  H  K  H  L  E  E  L  G  G  L  G     79
ACTGTGGCTGCTGGAGCCTTTGCACTCCACGAGAAGCACGCATCGAAGAAGGATCCAGAG   360
 T  V  A  A  G  A  F  A  L  H  E  K  H  A  S  K  K  D  P  E     99
AACGCTCACAGGCACAAGATTGAGGAGGAGATAGCTGCAGCAGCTGCAGTGGGAGCAGGG   420
 N  A  H  R  H  K  I  E  E  E  I  A  A  A  A  A  V  G  A  G    119
GGTTACGTGTTCCACGAGCATCACGAGAAGAAAGAATCGAAAGAAGAAGAAAAGGAAGCA   480
 G  Y  V  F  H  E  H  H  E  K  K  E  S  K  E  E  E  K  E  A    139
GAGGGCAAAAAGCATCACCACCTCTTCTACGTGCGCTGTCACTAGTCTTGGCTTTAAAAT   540
 E  G  K  K  H  H  H  L  F  Y  V  R  C  H  *                  153
AATGTCCATCTCTTTCCACTGTTGGTTGAATTTGGTGTGGGGTTGTTATGAGATTGAGAT   600
GAAGCTGAAATGAAGTGATACCCAAGTGGTCTTCAATCTTCACCTCCTCCCCTTCTTGTG   660
TTTGTTATCTTAGTTGCGTATCTCACATCCGTTCTTGATGGGATCGTTGTGTACTGATCC   720
GAATAAATTTATGAATAGAGATTGTGTTGTGAATAATCTAAAAAAAAAAAAAAAAAAA    778
```

Figure 5.

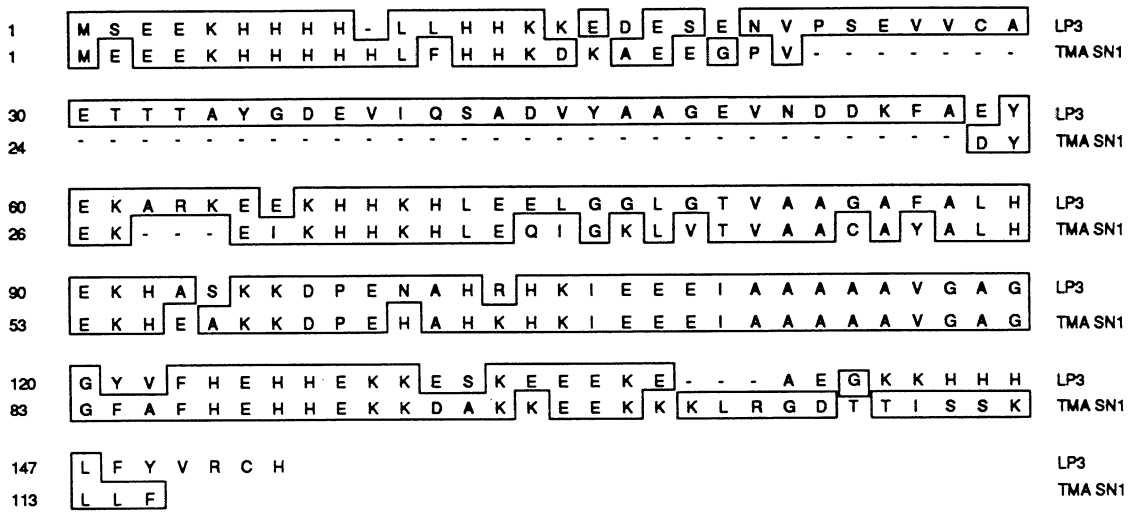
1 M S E E K H H H H - L L H H K K E D E S E N V P S E V V C A LP3
1 M E E E K H H H H H L F H H K D K A E E G P V - - - - - - - - TMA SN1
30 E T T T A Y G D E V I Q S A D V Y A A G E V N D D K F A E Y LP3
24 - D Y TMA SN1
60 E K A R K E E K H H K H L E E L G G L G T V A A G A F A L H LP3
26 E K - - - E I K H H K H L E Q I G K L V T V A A C A Y A L H TMA SN1
90 E K H A S K K D P E N A H R H K I E E E E I A A A A A A V G A G LP3
53 E K H E A K K D P E H A H K H K I E E E E I A A A A A A V G A G TMA SN1
120 G Y V F H E H H E K K E S K E E E K E - - - A E G K K H H H LP3
83 G F A F H E H H E K K D A K K E E K K K L R G D T T I S S K TMA SN1
147 L F Y V R C H LP3
113 L L F TMA SN1

Figure 6.

```
GGCACGAGCTGGTGTTCAAATACACGCCGGGCGCGCACAATACATTGGTGGTAAACAAGG      60
   H  E  L  V  F  K  Y  T  P  G  A  H  N  T  L  V  V  N  K       19
CGGCGTACGATGCGTGCACCCTTACTAATGCCTTGGCAACATACACCAGTGGCAACGACA     120
 A  A  Y  D  A  C  T  L  T  N  A  L  A  T  Y  T  S  G  N  D       39
CCATTTCGTTGAACAGCACGGGCGCCAAGTATTATATCTGCGGAATCCCAGGACACTGCT     180
 T  I  S  L  N  S  T  G  A  K  Y  Y  I  C  G  I  P  G  H  C       59
CCGGCGGCATGAAGCTGACTGTCACCGTTGCCGCCGCAAAGAGCAACGGAACGGCGCCAT     240
 S  G  G  M  K  L  T  V  T  V  A  A  A  K  S  N  G  T  A  P       79
CGCCCTCCCCCACTTCAAAGAGCAACGGGACGGCGCCATCGCCCTCCCCCTCCAGCTCTC     300
 S  P  S  P  T  S  K  S  N  G  T  A  P  S  P  S  P  S  S  S       99
CACCATCCCCCTCGCCCACCAGCTCTCCGCCATCCACCACTCCCACACCACCAATTGCCA     360
 P  P  S  P  S  P  T  S  S  P  P  S  T  T  P  T  P  P  I  A      119
CATCGCCTTCCCCATCTTCGGGCACTTCCCCCACTGGTACGTCTACAGGGTCACCGCCGC     420
 T  S  P  S  P  S  S  G  T  S  P  T  G  T  S  T  G  S  P  P      139
CGGAATCCACAACTACGCCTTCACCGAGTGGATCAAACAATAGCGCAGCCGCCCCTTCAT     480
 P  E  S  T  T  T  P  S  P  S  G  S  N  N  S  A  A  A  P  S      159
TCCGCTTGGATGGTGCCCTCCTCCTTGCAGGGTCACTTTTGCTGGCCATGGCGTCGCTAT     540
 F  R  L  D  G  A  L  L  L  A  G  S  L  L  L  A  M  A  S  L      179
AGGCATGCATAGATACCGAGCAATTATGTTGATTTCACTGAGTGAGCGCCTTTTTCACGA     600
 *
TTCCTGGGTGCATATATATTATTTTCTGCTCATCATGTACTATTGTGGTGATCATCGTTG     660
ACCGGATGCATTGGGTTCATGTCCGTCACGCTAATAAAATGGATTCATATGGTTACTTTA     720
TAAAAAAAAAAAAAAAAAAAAAAAA                                       745
```

Figure 7.

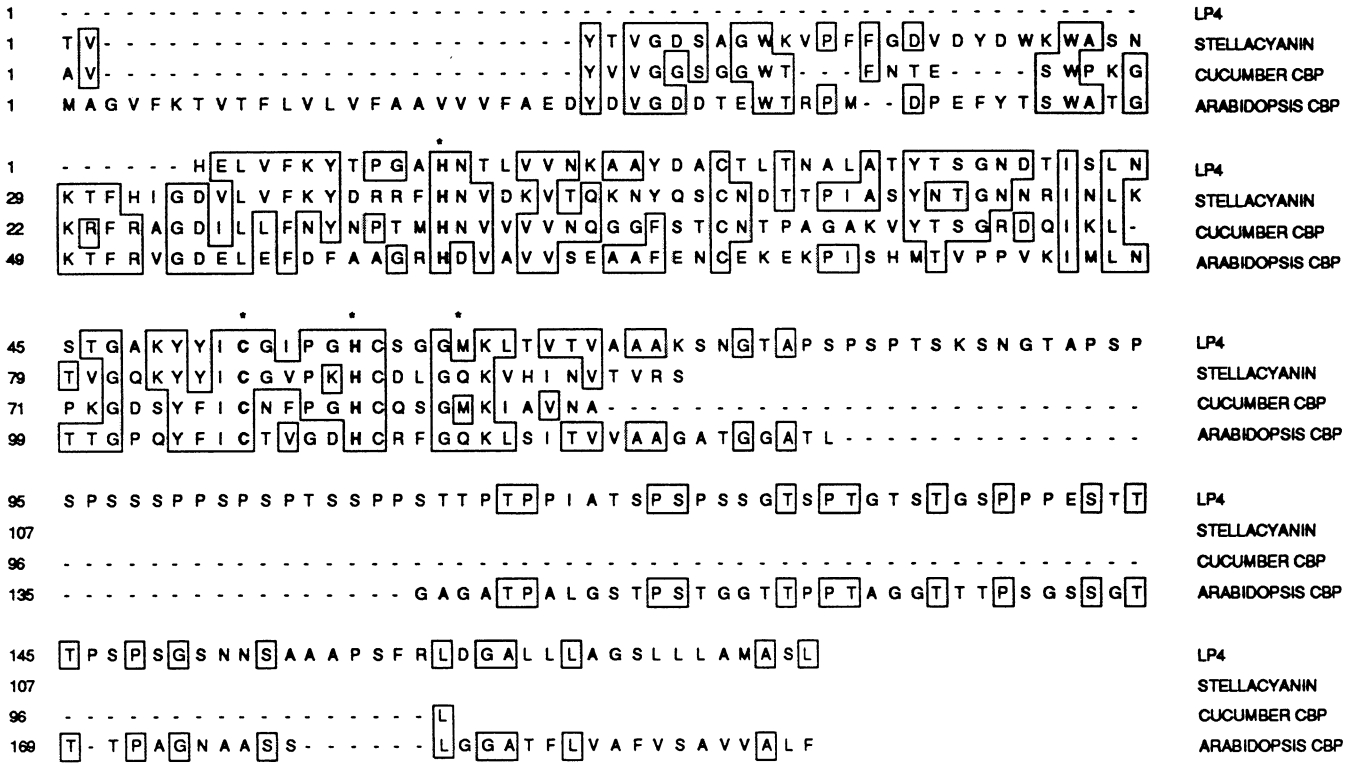

Figure 8.

```
GGCACGAGGTTGCATATACTTTGAACTAAGGAGAAGGAAAGGGGTGGTGAGAATTTGCGA     60
AGCAATGTCGAAGCAAAAACTCTTAATCTTTGCCGCCATGGCCGGACTTCTGTTTGCTTG    120
                                        M  A  G  L  L  F  A  C      8
TGCCGCCGTTGAGAGCCGTATTGCACGAAGCGACCTGGGCCTCGACCTCGGAGGAGGTTT    180
  A  A  V  E  S  R  I  A  R  S  D  L  G  L  D  L  G  G  G  L     28
GGGTCTCGGCGTCGGCGTTGGTGCAGGCTTGGGTTTGGGTGGCGGCAGCGCGAGCGGTAG    240
  G  L  G  V  G  V  G  A  G  L  G  L  G  G  G  S  A  S  G  S     48
CGGTTCTGGTTCTGGCTCCGGCTCTGGCTCTGGTTCCGGTGCCGGTTCCGCCGCTGGCTC    300
  G  S  G  S  G  S  G  S  G  S  G  S  G  A  G  S  A  A  G  S     68
AGGTTCCGGTTCTGGAGCTGGCTCCGGCGCTGGGTCTTACGCTGGCTCAGGAGCAGGCAA    360
  G  S  G  S  G  A  G  S  G  A  G  S  Y  A  G  S  G  A  G  N     88
CGGAGGCGGCCAGGGTCGTGGTTCCGGAAGTGGCTACGGTTCGGGTAGCGGTTACGGAGC    420
  G  G  G  Q  G  R  G  S  G  S  G  Y  G  S  G  S  G  Y  G  A    108
TGGTAATGGCAACGGCAATGGCTACGGCGCGGGTAGCGGTTACGGAGCTGGTAATGGCAA    480
  G  N  G  N  G  N  G  Y  G  A  G  S  G  Y  G  A  G  N  G  N    128
CGGCAATGGCTACGGCGCGGGTAGCGGCTCGGGTTCAGGCAGTGGTTCTGGCAGCGGCAG    540
  G  N  G  Y  G  A  G  S  G  S  G  S  G  S  G  S  G  S  G  R    148
GGGCTACGGTAGTGGCTCCGGTACTGGTTCGGGCTACGGTTCCGGTTCCGGTTCTGGTTA    600
  G  Y  G  S  G  S  G  T  G  S  G  Y  G  S  G  S  G  S  G  Y    168
CGGTAACGGTTCTGGTTCTGGTTCCGGTTACGGCGCTGGAGATGATGGCAGCAACGAGGG    660
  G  N  G  S  G  S  G  S  G  Y  G  A  G  D  D  G  S  N  E  G    188
AGCTTCTGGTGGCGGATATTAAATCTATAAAAAAAACTAAAATAGTATGCGTAAAACCGTT    720
  A  S  G  G  G  Y  *                                           194
ATTATCATTTGCATGGTGTTGAATCTATTGTACTAAACAAATAAACAGGCCATGGCATCT    780
CTTTAGCATTTCCTTCATCTCTGCACAAACTGATCATTTCTCTTTAGTTCATCAGATCCA    840
GATTTGAGATATCTACAGTGTATTGTGTGGCCATGGTGTGCTTCTTAAATTTAGCTTTCC    900
TTCATTAAGTGTTAGCCAATGTAAGATTTCTGCAAATATTAAGAAGAGCTATTGTAGAAG    960
TAGTCATAAACAAAAAAAAAAAAAAAA                                     988
```

Figure 9.

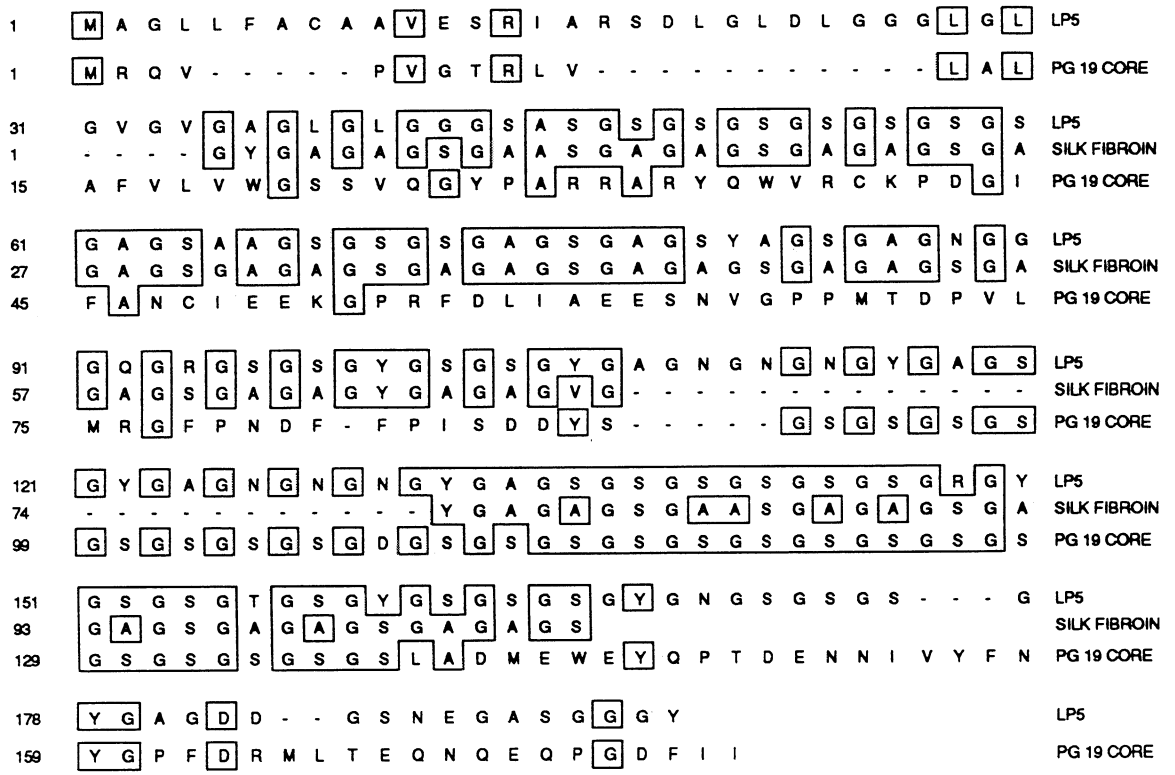
1 M A G L L F A C A A V E S R I A R S D L G L D L G G G L G L LP5
1 M R Q V - - - - - P V G T R L V - - - - - - - - - - - - L A L PG 19 CORE
31 G V G V G A G L G L G G G S A S G S G S G S G S G S G S G S LP5
1 - - - - G Y G A G A G S G A A S G A G A G S G A G A G S G A SILK FIBROIN
15 A F V L V W G S S V Q G Y P A R R A R Y Q W V R C K P D G I PG 19 CORE
61 G A G S A A G S G S G S G A G S G A G S Y A G S G A G N G G LP5
27 G A G S G A G A G S G A G A G S G A G A G S G A G A G S G A SILK FIBROIN
45 F A N C I E E K G P R F D L I A E E S N V G P P M T D P V L PG 19 CORE
91 G Q G R G S G S G Y G S G S G Y G A G N G N G N G Y G A G S LP5
57 G A G S G A G A G Y G A G A G V G - - - - - - - - - - - - - SILK FIBROIN
75 M R G F P N D F - F P I S D D Y S - - - - - G S G S G S G S PG 19 CORE
121 G Y G A G N G N G N G Y G A G S G S G S G S G S G S G R G Y LP5
74 - - - - - - - - - - - Y G A G A G S G A A S G A G A G S G A SILK FIBROIN
99 G S G S G S G S G D G S G S G S G S G S G S G S G S G S G S PG 19 CORE
151 G S G S G T G S G Y G S G S G S G Y G N G S G S G S - - - G LP5
93 G A G S G A G A G S G A G A G S SILK FIBROIN
129 G S G S G S G S G S L A D M E W E Y Q P T D E N N I V Y F N PG 19 CORE
178 Y G A G D D - - G S N E G A S G G G Y LP5
159 Y G P F D R M L T E Q N Q E Q P G D F I I PG 19 CORE

Figure 10.